INFORMATION AND CONTROL 8, 56–63 (1965)

Generalizations of Regular Events

V. AMAR AND G. PUTZOLU

Laboratorio Ricerche Elettroniche Olivetti Pregnana Milanese, Milano, Italy

INTRODUCTION

In a previous paper we have introduced a family of linear grammars, which we have called even linear grammars (ELG), and we have shown that the languages generated by the ELG's contain the regular events and behave analogously to them.

In the present paper we point out the mathematical reasons of the regular-like behavior of these languages, which enables us to obtain a sort of "Kleene Theorem" for them. We also generalize all results about ELG's to a class of families of linear grammars. The languages generated by each family have properties similar to regular events, and contain them. Finally an interesting problem, in our opinion, about the intersection of all these families is pointed out.

SECTION I

We briefly recall the main results of an earlier paper (Amar and Putzolu, 1964), hereafter denoted by PI.

Let $\Sigma = \{\sigma\}$ be a finite alphabet, which will be fixed for the rest of the discussion, and let $T_\Sigma = \{\varphi, \psi \cdots\}$ be the free semigroup with unity λ on Σ.

DEFINITION 1. An even linear grammar (ELG) on the alphabet Σ is a context free grammar (Chomsky and Schützenberger, 1963; Bar-Hillel, Perles, and Shamir, 1961; Ginsburg and Rice, 1962) $\mathcal{G} = (\Delta, \delta_0, P)$ where Δ is the auxiliary alphabet, δ_0 is the initial symbol, and P has only productions of the form:

$$\begin{cases}\delta \to \varphi \\ \delta \to \varphi\bar{\delta}\psi\end{cases} \qquad \text{with}^1 \quad |\varphi| = |\psi|.$$

The languages generated by the ELG's are called quasi-regular events.

[1] $|\varphi|$ denotes the length of the word φ.

DEFINITION 2. An equivalence relation E on T_Σ is called a quasi congruence if it satisfies the following condition:

$$\text{if} \quad \varphi E \psi \qquad \text{then} \quad \begin{cases} \sigma' \varphi \sigma'' E \sigma' \psi \sigma'' \\ \forall (\sigma', \sigma'') \end{cases}$$

THEOREM 5 (PI). *A set H is a quasi-regular event if and only if there exists a quasi congruence of finite index which saturates it.*

Since every congruence is a quasi congruence, from the theorem of Myhill (Rabin and Scott, 1959) we have that quasi-regular events contain regular events.

SECTION II

We now illustrate the mathematical reasons of the regular-like behavior of quasi-regular events.

Given a word $\varphi \in T_\Sigma$ we decompose it in this way:

$$\varphi = \varphi' \omega_\varphi \varphi'' \quad \text{where} \quad \begin{cases} |\varphi'| = |\varphi''| \\ \omega_\varphi \in \Sigma \cup \lambda = \Omega \end{cases}$$

This decomposition is clearly unique. We define then on T_Σ the following abstract product denoted by "$\circ$":

$$\varphi \circ \psi \overset{\text{def}}{\equiv} \psi' \varphi \psi'' = \psi' \varphi' \omega_\varphi \varphi'' \psi''$$

II. 1. *"$\circ$" is associative. In fact:*

$$\varphi \circ (\psi \circ \chi) = \varphi \circ (\chi' \psi' \omega_\psi \psi'' \chi'') = \chi' \psi' \varphi \psi'' \chi''$$
$$= (\psi' \varphi \psi'') \circ \chi = (\varphi \circ \psi) \circ \chi$$

This implies that T_Σ is a semigroup with respect to the product "$\circ$". We shall call it $T^{(\circ)}$. It may be easily seen that $T^{(\circ)}$ has all the words $\varphi \in \Omega$ (i.e. $|\varphi| < 2$) as right unities, while it does not have any left unity. Furthermore $T^{(\circ)}$ is right cancellative (i.e., $\varphi \circ \psi = \chi \circ \psi$ implies $\varphi = \chi$) but not left cancellative.

II. 2. *Every right invariant equivalence relation on $T^{(\circ)}$ is a quasi congruence on T_Σ and conversely.*

PROOF: Let E be a right invariant equivalence relation on $T^{(\circ)}$. Then:

$$\varphi E \psi \rightarrow \begin{cases} \forall \chi \\ \varphi \circ \chi E \psi \circ \chi \end{cases} \rightarrow \begin{cases} \forall (\chi', \chi'') \quad \text{with} \quad |\chi'| = |\chi''| \\ \chi' \varphi \chi'' E \chi' \psi \chi'' \end{cases}$$

i.e., E is a quasi congruence; analogously we can prove the converse.

Using statement II. 2 it is now obvious that theorem 5 (PI) assumes a form under which it is strictly analogous to the theorem of Nerode (Rabin and Scott, 1959) for finite automata, i.e.:

II. 3. *Each quasi-regular event is saturated by a right invariant (with respect to the product "$\circ$") equivalence relation of finite index, and conversely.*

It now becomes clear why quasi-regular events have mathematical properties very similar to those of regular events. In fact both families are associated in the same way with right invariant equivalence relations respectively on $T^{(\circ)}$ and T_Σ. This remark clarifies that the research of families of events, which behave analogously to regular events, is essentially the research of suitable abstract product on the set of tapes.

It seems interesting therefore to study the intrinsic properties of semi-groups defined in this way, and to try establishing relations between these properties and the features of the classes of the associated languages. The families of languages so generated are not necessarily context free but ought to have some relevant mathematical structure.

In Section IV we shall introduce a set of families of context free languages similar to regular languages and we shall show that each of these families is associated with a particular abstract product on the set of tapes.

SECTION III

We shall now study the closure properties of quasi-regular languages.

A. Complementation

If A is a quasi-regular event, there exists a quasi congruence of finite index E which saturates A (theorem 5 (PI)). Consequently E saturates also the complement of A in T_Σ, $\bar{A}$, i.e., quasi-regular events are closed under complementation.

B. Union and Intersection

If A and B are quasi-regular events, there exist two quasi congruences E_A and E_B of finite index which saturate respectively A and B.

It is easy to show that $E_A \cap E_B$ (i.e., the equivalence relation having as classes of intersection of the classes of E_A and E_B) is a quasi congruence and saturates $A \cup B$ and $A \cap B$. Consequently quasi-regular events are closed with respect to union and intersection.

In order to obtain other remarkable closure properties we give the following definitions:

1. If A and B are two sets of strings, then $A \circ B$ denotes the set of all elements of the form $\varphi \circ \psi$ with $\varphi \in A$, $\psi \in B$.

2. If A is a set of strings, then we define

$$A^{\dagger} = \Omega + A + A \circ A + \cdots$$ [2]

C. Operation "$\circ$"

Let A and B be two quasi-regular events, respectively generated by the ELG's $\mathcal{G}^A \equiv (\Delta^A, \delta_0^A, P^A)$ and $\mathcal{G}^B = (\Delta^B, \delta_0^B, P^B)$. We define a new grammar $\mathcal{G} = (\Delta, \delta_0, P)$ where $\Delta = \Delta^A + \Delta^B$, $\delta_0 = \delta_0^B$ and P is formed by these rules:

1. All the rules of P^A.
2. The rules of P^B not of the form $\delta^B \to \varphi$ (terminating rules).
3. $\delta^B \to \varphi' \delta_0^A \varphi''$ for each rule of P^B of the form $\delta^B \to \varphi = \varphi' \omega_\varphi \varphi''$. G is an ELG, and clearly $\mathcal{L}(\mathcal{G}) = A \circ B$, i.e., quasi-regular events are closed under the "$\circ$" operation.

D. Operation "$\dagger$"

Let A be a quasi-regular event generated by the ELG $\mathcal{G}^A = (\Delta^A, \delta_0^A, P^A)$. We define a new grammar $\mathcal{G} = (\Delta, \delta_0, P)$ where $\Delta = \Delta^A$, $\delta_0 = \delta_0^A$ and P contains all the rules of P^A, the rules $\delta_0 \to \omega$ for each $\omega \in \Omega$, and also a rule $\delta^A \to \varphi' \delta_0 \varphi''$ for each rule $\delta^A \to \varphi = \varphi' \omega_\varphi \varphi''$ of P^A. $\mathcal{G}$ is an ELG, and clearly $\mathcal{L}(\mathcal{G}) = A^{\dagger}$, i.e., quasi-regular events are closed under the "$\dagger$" operation.

We now give a theorem, which we call the "Kleene theorem" for the family of quasi-regular events.

Theorem 1. *The family of quasi-regular events is the minimal family of sets of tapes which contains the finite sets, and which is closed under* $+$, "$\circ$," *and* "$\dagger$."

To prove the theorem we need the following lemma.

III. (Brzozowski, 1962). *Let A and B be two sets, with the condition that for each $\psi \in A$ it is $|\psi| \geqq 2$. In other words A does not contain any right unity of $T^{(\circ)}$. Consider the formal equation on sets*

$$X = X \circ A + B \tag{1}$$

[2] Here and in the following we use $+$ to indicate the logical union $\cup$.

We prove that its solution is:

$$X = B \circ A^\dagger$$

1. $X \supset B \circ A^\dagger$. In fact, from (1) we have that $X \supset B$ and $X \supset X \circ A$. Suppose inductively that $X \supset B \circ A^n$ which is true for $n = O(X \supset B)$. We have $B \circ A^{n+1} = B \circ A^n \circ A \subset X \circ A \subset X$.

2. $X \subset B \circ A^\dagger$. Let us suppose that, contrary to the assumptions, there exists a φ such that $\varphi \in X$, $\varphi \notin B \circ A\dagger$. From (1) there must exist a pair (φ_1, ψ_1) with $\varphi_1 \in X$, $\psi_1 \in A$ such that $\varphi = \varphi_1 \circ \psi_1$. Since $|\psi_1| \geqq 2$ we have that $|\varphi| > |\varphi_1|$. From the $\varphi_1 \in X$ it follows that $\varphi_1 \in X \circ A$ or $\varphi_1 \in B$. But $\varphi_1 \in B$ is against the hypothesis, so that there must exist a $\varphi_2 \in X$ and a $\psi_2 \in A$ such that $\varphi_1 = \varphi_2 \circ \psi_2$ with $|\varphi_2| < |\varphi_1|$. By a finite number of the above described steps we must find $\varphi = \varphi_n \circ \psi_n \circ \psi_{n-1} \circ \cdots \circ \psi_1$ with $\psi_i \in A (i = 1 \cdots n)$ and $\varphi_n \in B$, which is contrary to the hypothesis. Consequently $X \subset B \circ A^\dagger$.

Consider now an ELG $\mathcal{G} = (\Delta, \delta_0, P)$ which we may assume to be free of productions of the form $\delta_i \rightarrow \delta_j$. To each symbol δ_i we associate the formal equation

$$\delta_i = \sum_{p,q,r} \varphi_p' \delta_q \varphi_r'' + \sum_s \varphi_s = \sum_{r,q,r} \delta_q \circ (\varphi_p' \varphi_r'') + \sum_s \varphi_s$$

in the conventional way (Chomsky and Schützenberger, 1963; Ginsburg and Rice, 1962).

We have now a system of a finite number of formal equations in the sets δ_i. Solving the system with respect to δ_0, using repeatedly III. 1, we find that $\mathcal{L}(\mathcal{G})$ may be written applying the operators $+$, $\circ$, and $\dagger$ a finite number of times to a finite set of strings, q.e.d. We give a simple example.

Let us consider the grammar $\mathcal{G} = (\Delta, \delta_0, P)$ on $\Sigma = (0, 1)$, where:

$$\Delta = \{\delta_0, \delta_1\}$$

$$P = \begin{Bmatrix} \delta_0 \rightarrow 10\ \delta_0\ 10 \\ \delta_0 \rightarrow 0\ \delta_1\ 1 \\ \delta_1 \rightarrow 1 \\ \delta_1 \rightarrow 0\ \delta_1\ 0 \\ \delta_1 \rightarrow 1\ \delta_0\ 1 \end{Bmatrix}$$

We obtain the two formal equations:

$$\delta_0 = 10\delta_0 10 + 0\delta_1 1 = \delta_0 \circ (1010) + \delta_1 \circ (01)$$

$$\delta_1 = 1 + 0\delta_1 0 + 1\delta_0 1 = 1 + \delta_1 \circ (00) + \delta_0 \circ (11)$$

Solving them, using III. 1 twice, we have:

$$\mathcal{L}(\mathcal{G}) = 1 \circ (00)^{\dagger} \circ (01) \circ \{(11) \circ (00)^{\dagger} \circ (01) + 1010\}^{\dagger}$$

SECTION IV

We shall now extend the results obtained about the family of the ELG's to a set of families of linear grammars.

Let $k = k'/k''$ be a rational number, k' and k'' being relatively prime.

Definition 3. A k-linear grammar is a context-free grammar having only productions of the form:

$$\begin{cases} \delta \to \varphi \\ \delta \to \varphi_1 \bar{\delta} \varphi_2 \quad \text{with} \quad |\varphi_1|/|\varphi_2| = k \end{cases}$$

We call k-regular events the sets generated by the family of k-linear grammars. According to Definition 3 the ELG's are the 1-linear grammars, and the quasi-regular events are the 1-regular events.

Definition 4. An equivalence relation on T_Σ is called a k-congruence if it satisfies the following condition:

$$\text{if} \quad \varphi E \psi \quad \text{then} \quad \begin{cases} \chi' \varphi \chi'' E \chi' \psi \chi'' \quad \text{with} \quad |\chi'|/|\chi''| = k \\ \forall (\chi', \chi'') \end{cases}$$

According to Definitions 2 and 4 the quasi congruences are the 1-congruences.

With techniques completely analogous to those of (PI), it is possible to show the following theorem:

Theorem 2. *A set H is a k-regular event if and only if there exists a k-congruence of finite index which saturates it.*

Given a word φ and a rational number $k = k'/k''$ (k', k'' relatively prime) we decompose φ in this way:

$$\varphi = \varphi_1 \varphi_0 \varphi_2 \quad \text{with} \quad \begin{cases} \dfrac{|\varphi_1|}{|\varphi_2|} = k \\ |\varphi_0| < k' + k'' \end{cases}$$

This decomposition is unique. In fact since $|\varphi_1|/|\varphi_2| = k'/k''$ it follows that $|\varphi_1| = mk'$, $|\varphi_2| = mk''$, with m integer. But:

$$|\varphi| = |\varphi_1| + |\varphi_2| + |\varphi_0| = m(k' + k'') + |\varphi_0|$$

$$|\varphi_0| < k' + k''$$

From the division algorithm follows the uniqueness of m and $|\varphi_0|$ and consequently of the given decomposition.

The meaning of Theorem 2 is now better clarified if we note that for each k we may introduce on T_Σ the following associative abstract product "$\circ_k$":

$$\varphi \circ_k \psi = \psi_1 \varphi \psi_2$$

In this way we induce on T_Σ a new semigroup structure with respect to "$\circ_k$". We call this semigroup $T_\Sigma^{(\circ_k)}$. It is easily verified that $T_\Sigma^{(\circ_k)}$ has no left unity, while it has each φ with $|\varphi| < k' + k''$ as right unity. We call Ω_k the set of all these unities. In addition $T_\Sigma^{(\circ_k)}$ is right but not left cancellative.

Observing that the k-congruences are just right invariant equivalence relations with respect to "$\circ_k$", Theorem 2 is the analogous of the theorem of Nerode with respect to this abstract product.

It is also possible to prove a Kleene Theorem for the family of k-linear grammars, involving the operators $+$, $\circ_k$, and $^{\dagger}{}_k$.[3]

$$(A^{\dagger k} = \Omega_k + A + A \circ_k A + \cdots)$$

We point out finally that since every congruence is also a k-congruence for each k, we have from Myhill's (Rabin and Scott, 1959) theorem this nonobvious result:

Given a one-sided (Chomsky and Schützenberger, 1963; Bar-Hillel *et al.*, 1961) linear grammar for each k there exists a k-linear grammar equivalent to it. This result allows us to define some strange regularity preserving operations on regular sets (Stearns and Hartmanis, 1963). Given a regular event R and $k = k'/k''$, consider the set

$$R_\alpha^{\ 1} = \{\varphi_1 \mid \exists \varphi = \varphi_1 \varphi_0 \varphi_2 \in R \ [|\varphi_1|/|\varphi_2| = k, \qquad |\varphi_0| = \alpha < k' + k'']\}$$

(α fixed)

We shall show that $R_\alpha^{\ 1}$ is regular. In fact let $\mathcal{G}$ be a k-linear grammar which generates R, construct the one sided linear grammar $\mathcal{G}'$ with the following conditions:

1. For any rule $\delta \to \varphi' \bar{\delta} \varphi''$ of $\mathcal{G}$ substitute $\delta \to \varphi' \bar{\delta}$.
2. For any rule $\delta \to \varphi = \varphi_1 \varphi_0 \varphi_2$ substitute $\delta \to \varphi_1$ if $|\varphi_0| = \alpha$. Otherwise delete the rule. $\mathcal{G}'$ generates $R_\alpha^{\ 1}$.

[3] The proof is based on the fact that the solution of the equation $X = X \circ_k A + B$ with the condition $|\psi| \geqq k' + k''$ for each $\psi \in A$, is $X = B \circ_k A^{\dagger k}$.

Theorem 3 also tells that the family of regular events is contained in the intersection of all the families of k-regular events. Do regular events exhaust this intersection? We have not been able to answer this question.

The problem seems to be interesting. In fact, in one case one would give a singular characterization of regular events, and in the other case one would find a new family of events, with pleasant properties: it would be properly contained in the families of k-regular events for each k, and it would properly contain regular events, being a sort of "minimal extension" of them.

RECEIVED: February 4, 1964.

REFERENCES

AMAR, V., AND PUTZOLU, G. (1964), On a family of linear grammars. *Inform. Control.* (to be referred to as (PI)).

BAR-HILLEL, Y., PERLES, M., AND SHAMIR, E. (1961), On formal properties of simple phrase structure grammars. *Z. Phonetic, Sprachwiss., Kommunikationsforsch.* **14,** 143–172.

BRZOZOWSKI, J. A. (1962), A survey of regular expressions and their applications. *IRE Trans. Electron. Computer* **EC-II,** 324–335.

CHOMSKY, N., AND SCHÜTZENBERGER, M. P. (1963), The algebraic theory of context-free languages. *In* "Studies in Logic—Computer Programming and Formal Systems." North Holland, Amsterdam.

GINSBURG, G., AND RICE, H. C. (1962), Two families of languages related to Algol. *J. Assoc. Computing Mach.* **9,** 350–371.

RABIN, M. O., AND SCOTT, D. (1959), Finite automata and their decision problems. *IBM J. Res.* **3,** 115–125.

STEARNS, R. E., AND HARTMANIS, J. (1963), Regularity preserving modifications of regular expressions. *Inform. Control* **6,** 55–69.